

# 实时数据仓库解决方案

美团网-鞠大升  
judasheng@meituan.com  
2015-11-03

# 大纲

- 案例：实时数据运营
- 实时计算平台
- 实时数据仓库
- 应用：RtDW直播间
- 经验总结
- 未来规划

# 大纲

## ➤ 案例：实时数据运营

➤ 实时计算平台

➤ 实时数据仓库

➤ 应用：RtDW直播间

➤ 经验总结

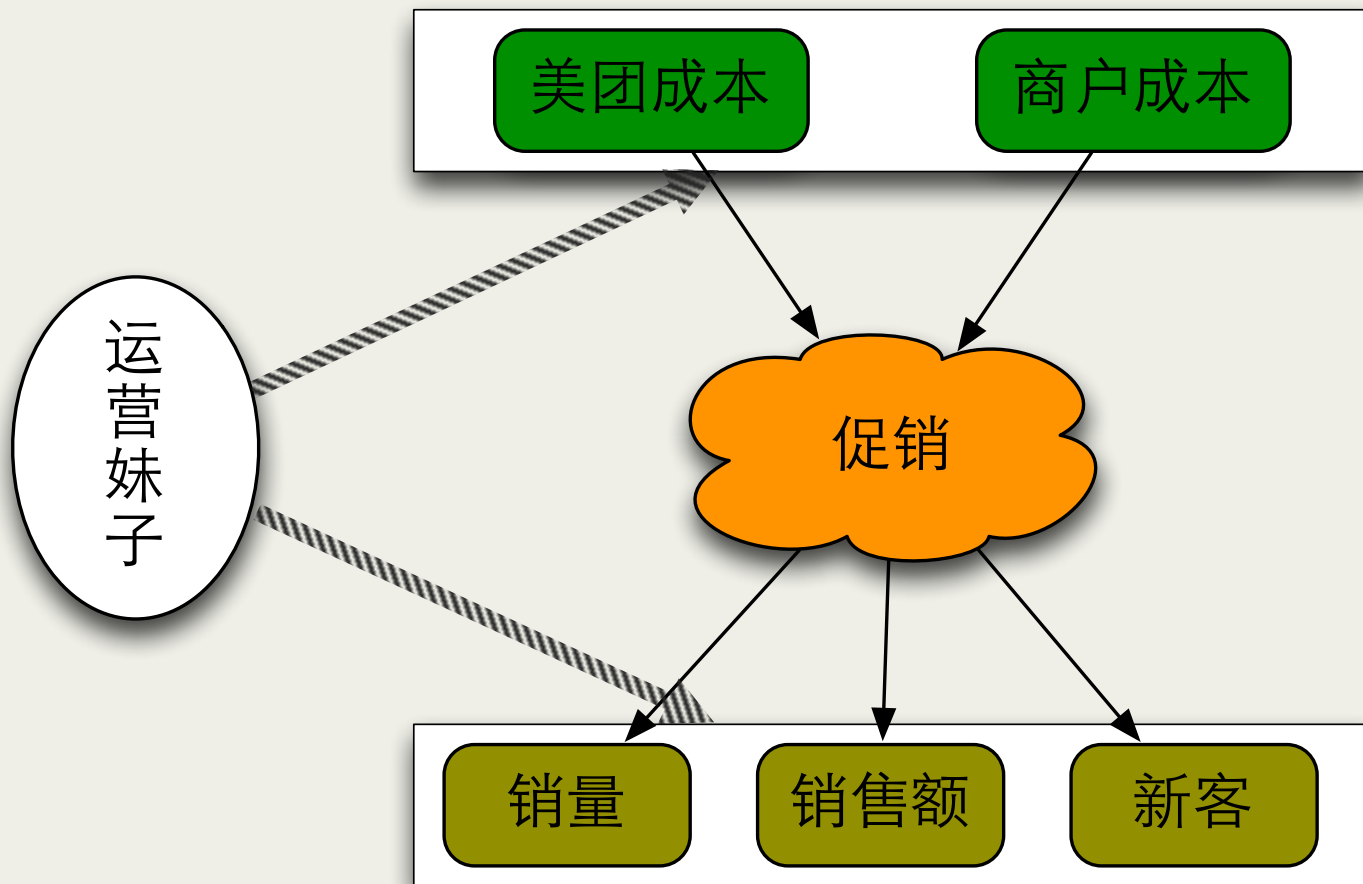
➤ 未来规划

# 什么是运营



- 运营就是一切帮助产品推广、使用、认知的手段。
- 三个元素：产品，用户，渠道；
- 运营目标：营收，扩大用户群，用户活跃度；
- 运营要关注成本和产出，ROI，投资回报率；
- 运营必须目标导向，数据为基础；

# 运营需求



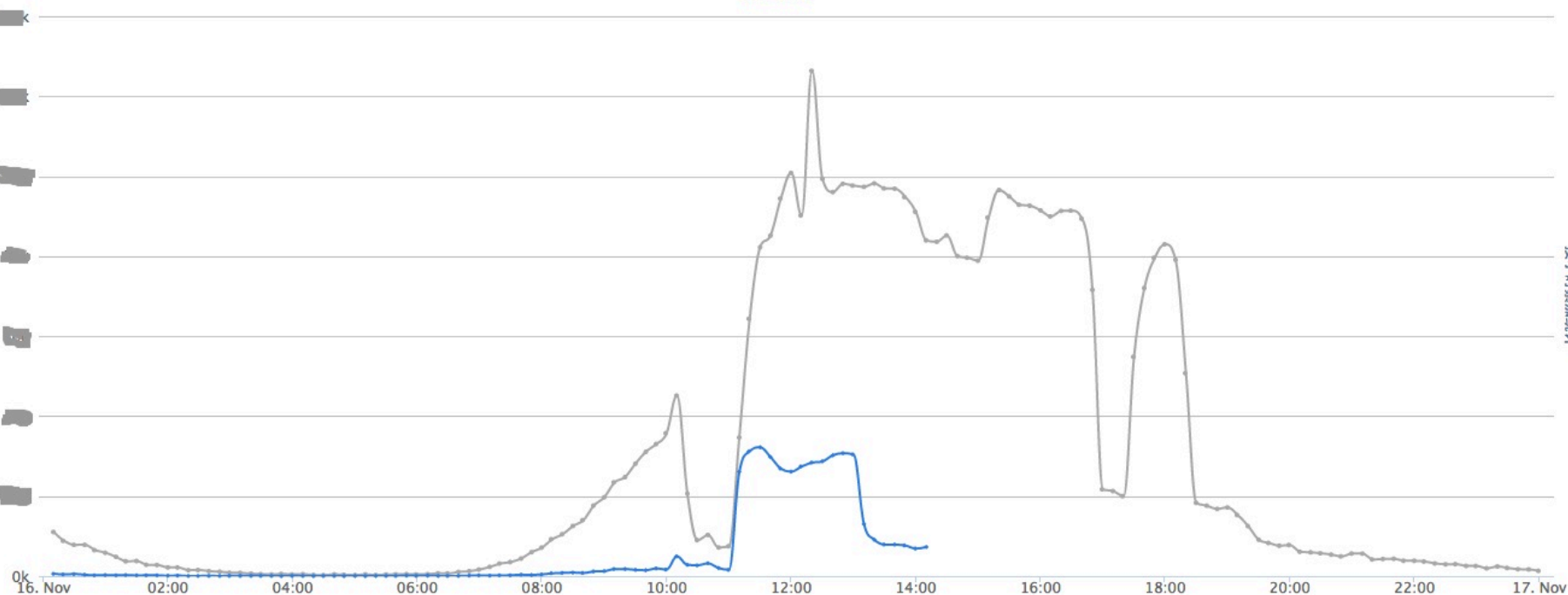
# 运营需求



猫眼总成本

基准: 今天 对比: 20151115 更新间隔: 10分钟 小时汇总: 不需要

猫眼平台总成本  
Download

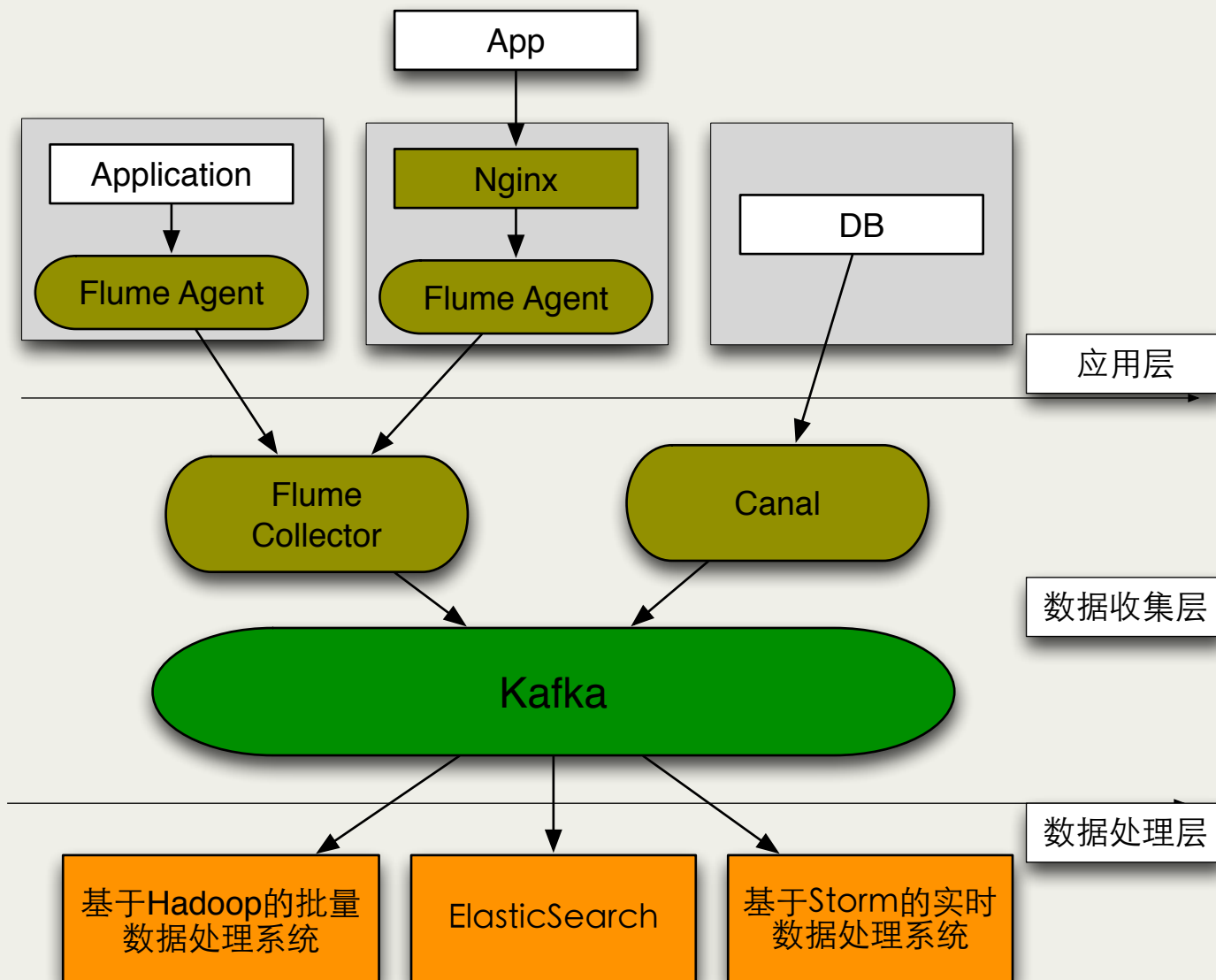


猫眼总成本\_20151115 猫眼总成本

# 大纲

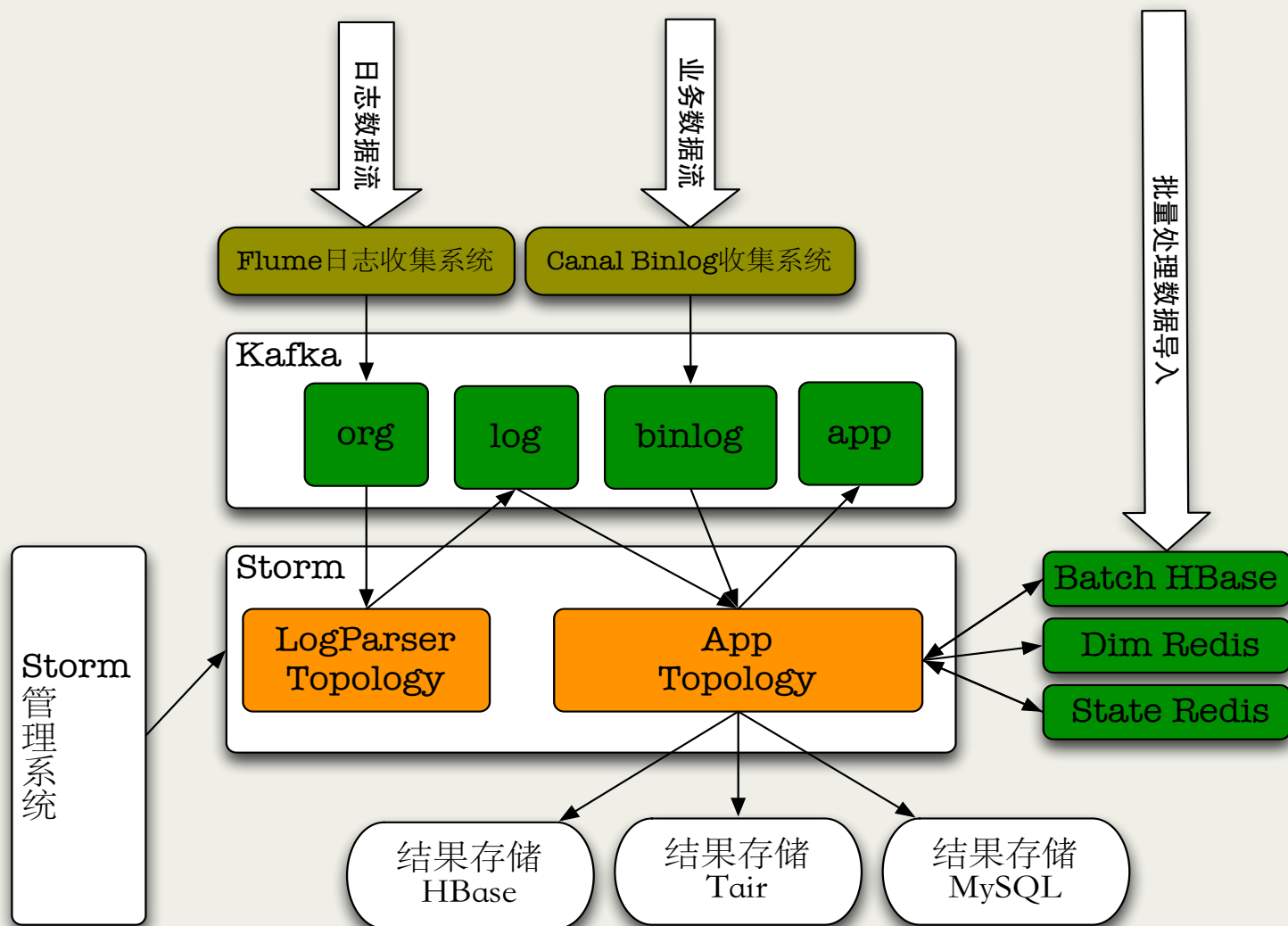
- 案例：实时数据运营
- 实时计算平台
- 实时数据仓库
- 应用：RtDW直播间
- 经验总结
- 未来规划

# 数据处理系统





# 实时计算平台



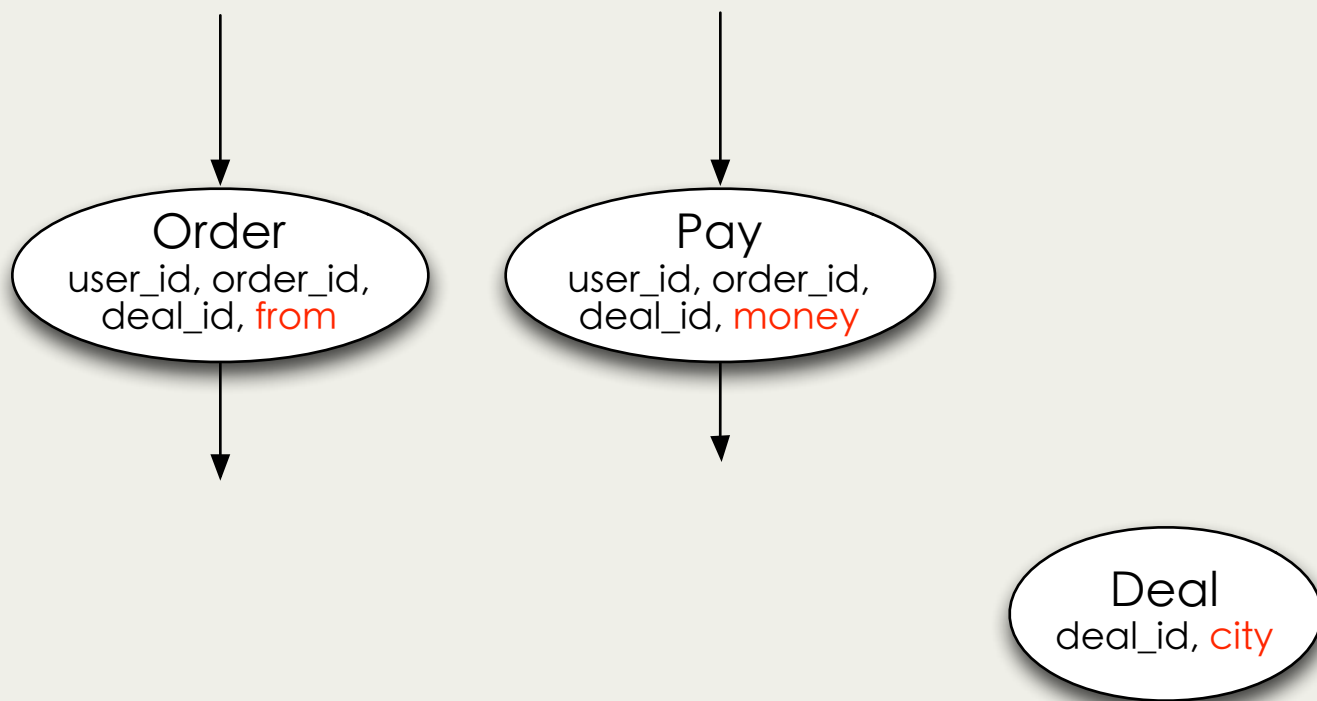
# 实时计算平台功能特性

- 开发
  - ✓ 测试开发平台;
  - ✓ 简单易用的拓扑框架;
  - ✓ *Input/Dim/State/Output*等一体化的解决方案;
- 服务
  - ✓ 延迟统计;
  - ✓ 依赖关系解析;
  - ✓ 报警服务;
  - ✓ 资源分组隔离;

# 大纲

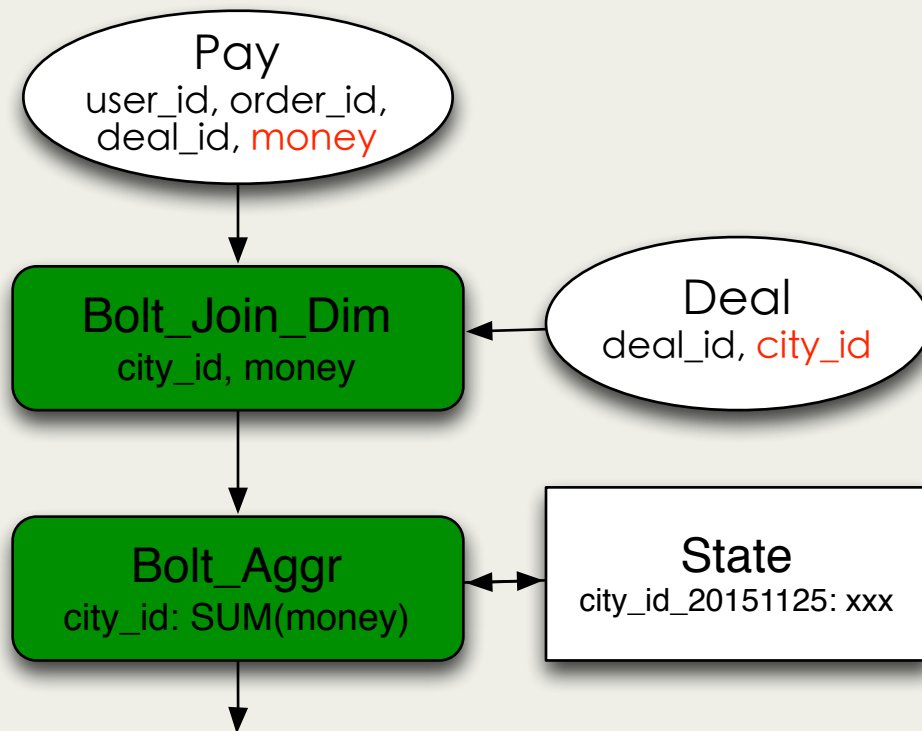
- 案例：实时数据运营
- 实时计算平台
- 实时数据仓库
- 应用：RtDW直播间
- 经验总结
- 未来规划

# 例子：需求场景

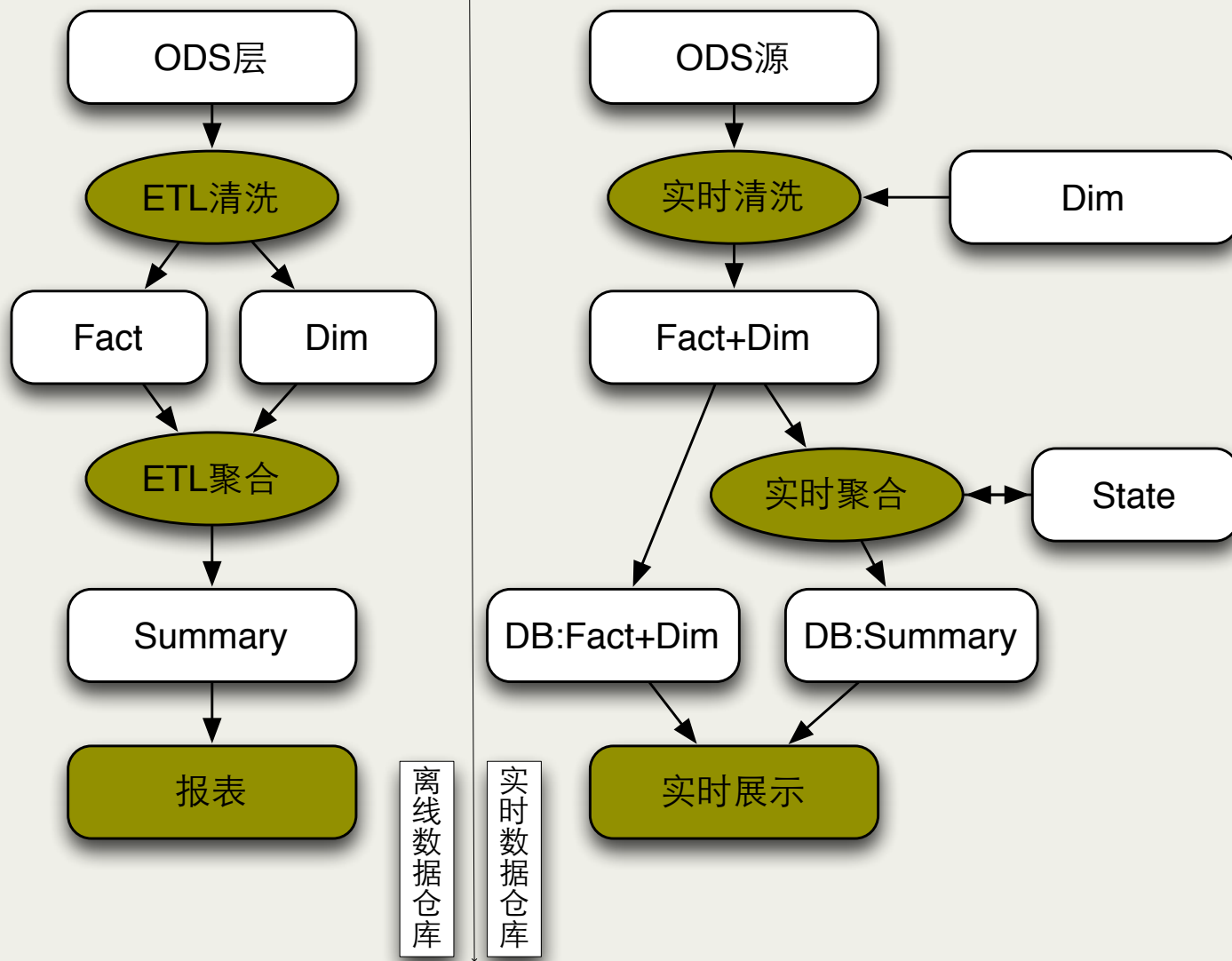


# 例子1

- 统计每个城市当天销售额



# 数据仓库架构

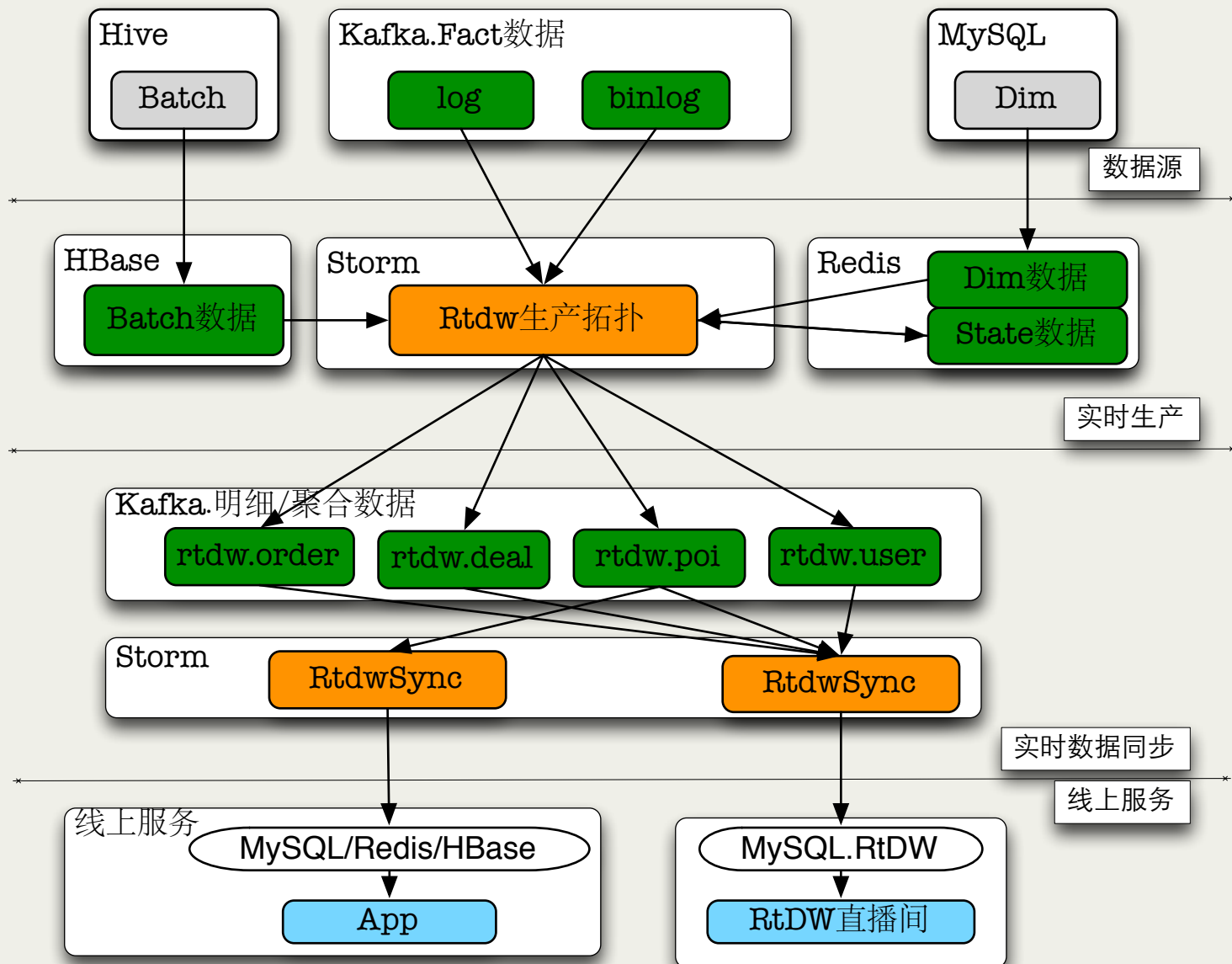


# 实时数据仓库区别



- 将事实 and 维度的Join前置到清洗阶段；
- 由于实时处理的低延时要求，维度Dim和状态需要在内存中访问；
- 由于当天的数据量有限，Fact+Dim可以存入DB中，支持展示时临时聚合；
- 实时数据展示区别于报表展示，需实时更新；

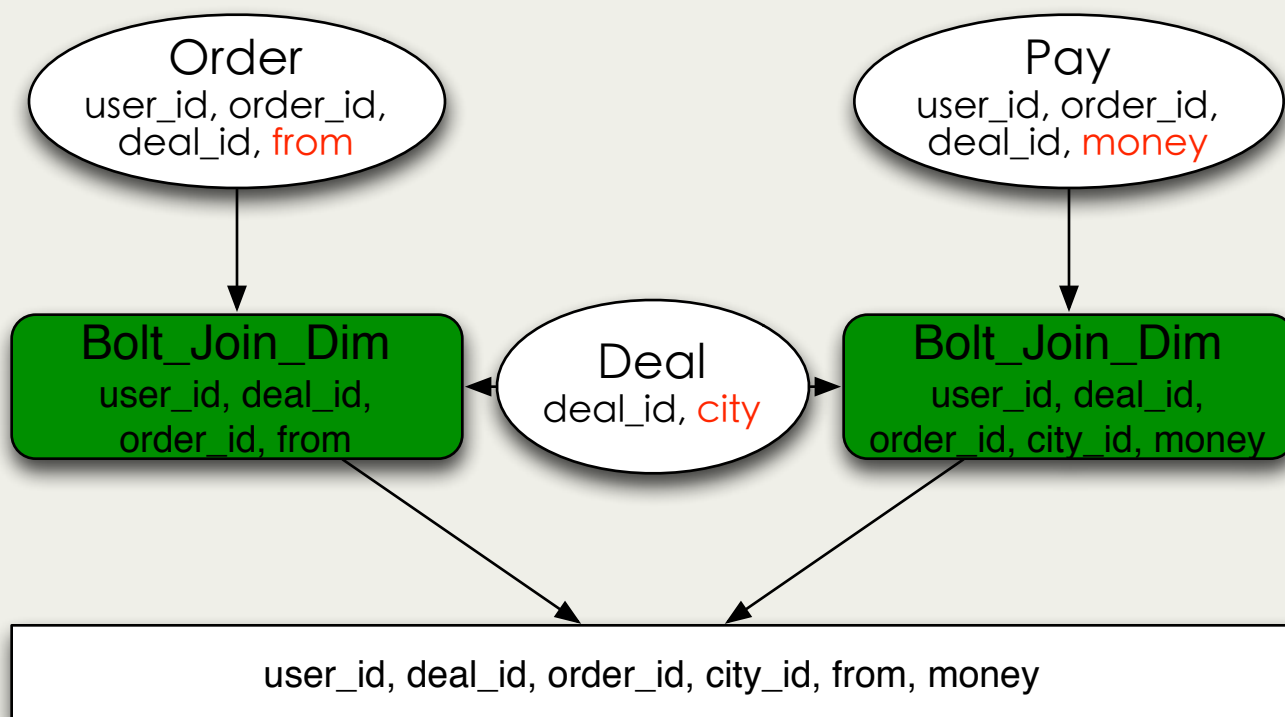
# 实时数据仓库





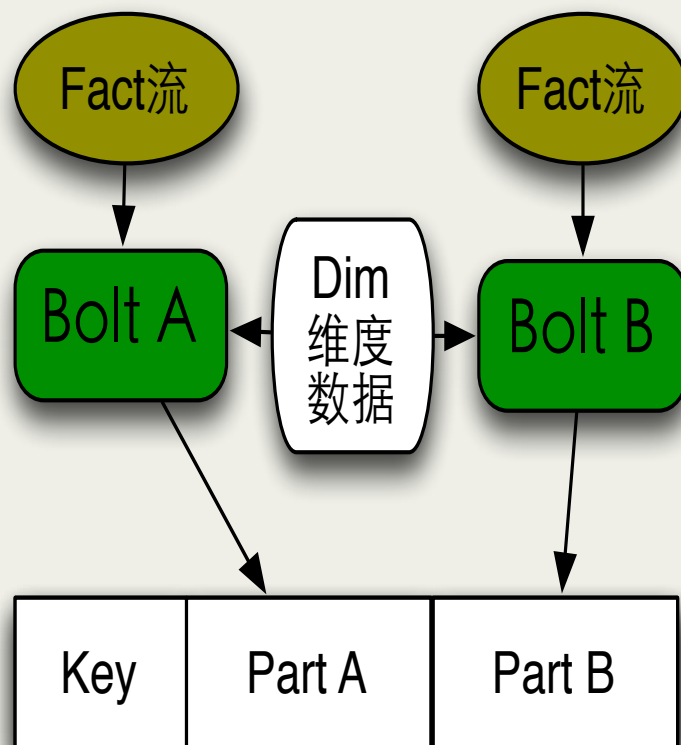
# 例子2

- 按Deal, From, City等维度统计当天销售额



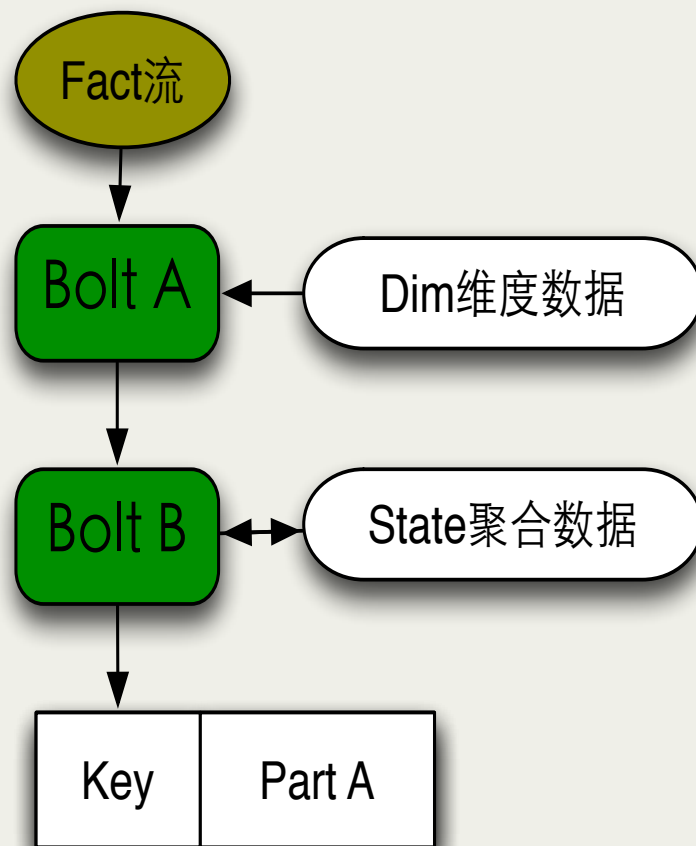
# 策略-明细数据生产方案

- ✓ 适合事实(Fact)数据清洗, 如订单数据;
- ✓ 通过DimRedis支持Join维度数据;
- ✓ 通过数据库, 支持多Fact流Join, 数据按天切表;
- ✓ 支持需要二次聚合, 如按时间段统计订单量;

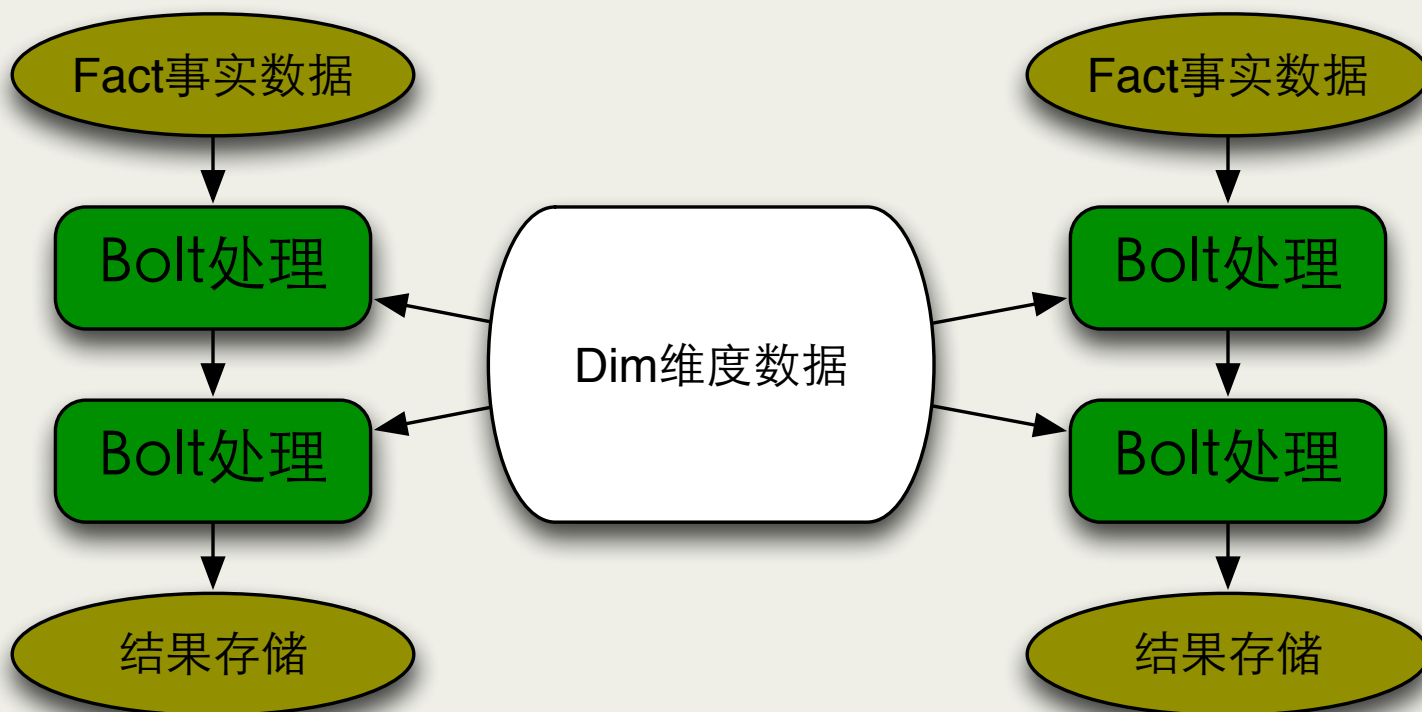


# 策略-聚合数据生产方案

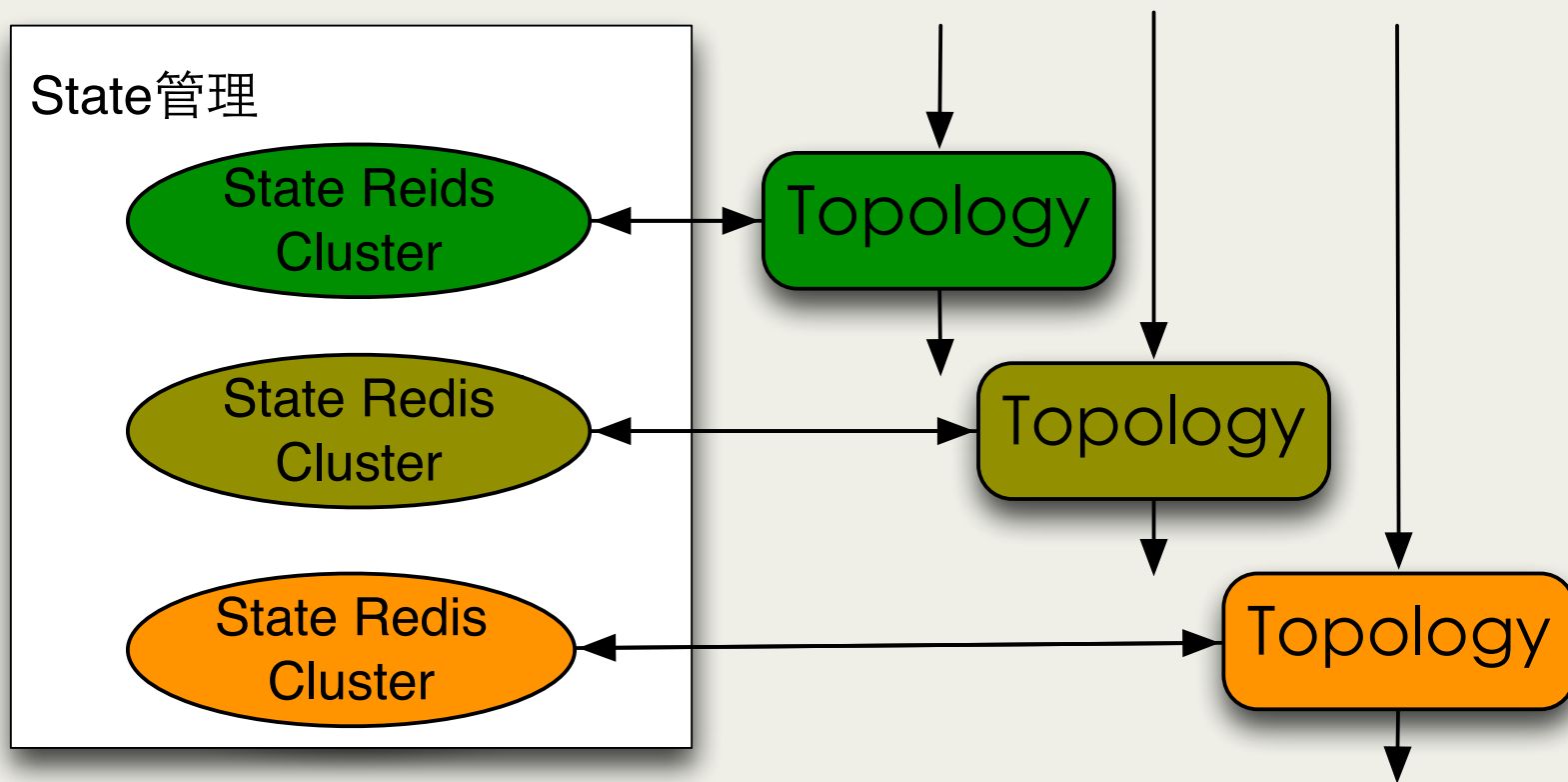
- ✓ 适合按维度统计，如按 Deal/POI 统计结果；
- ✓ 通过 DimRedis 支持 Join 维度数据；
- ✓ 利用 StateRedis 支持聚合；
- ✓ 将聚合最终结果写出，而不是增量结果；



# 策略-维度数据统一建设

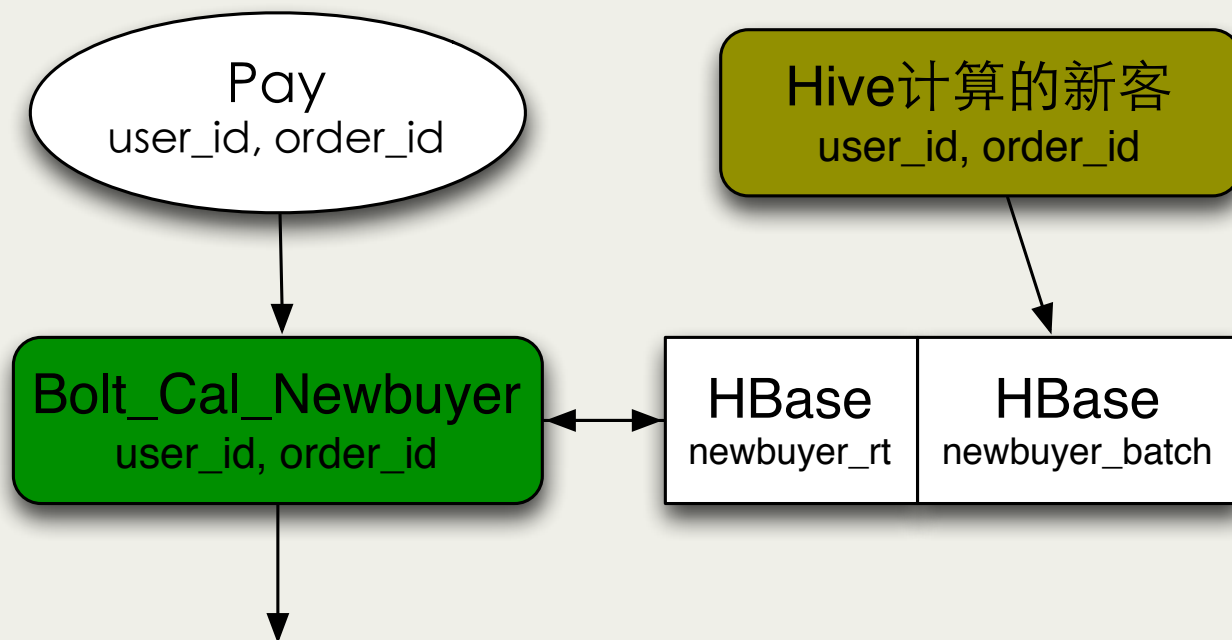


# 策略-状态数据存储



# 例子3

- 实时统计新客



- 小数据量去重：利用Redis State的Set功能；
  - 如：求Deal当天的UV；
- 大数据量精确去重：利用HBase的Key去重；
  - 如：美团的新客计算；
- 大数据量约数去重：利用Redis State的HyperLogLog算法；

# 实时数据仓库现状





# 实时数据仓库应用

- 大屏幕展示;
- 数据挖掘和筛选;
- 业务系统监控和预警;
- RtDW直播间: 实时数据化运营;

# 大纲

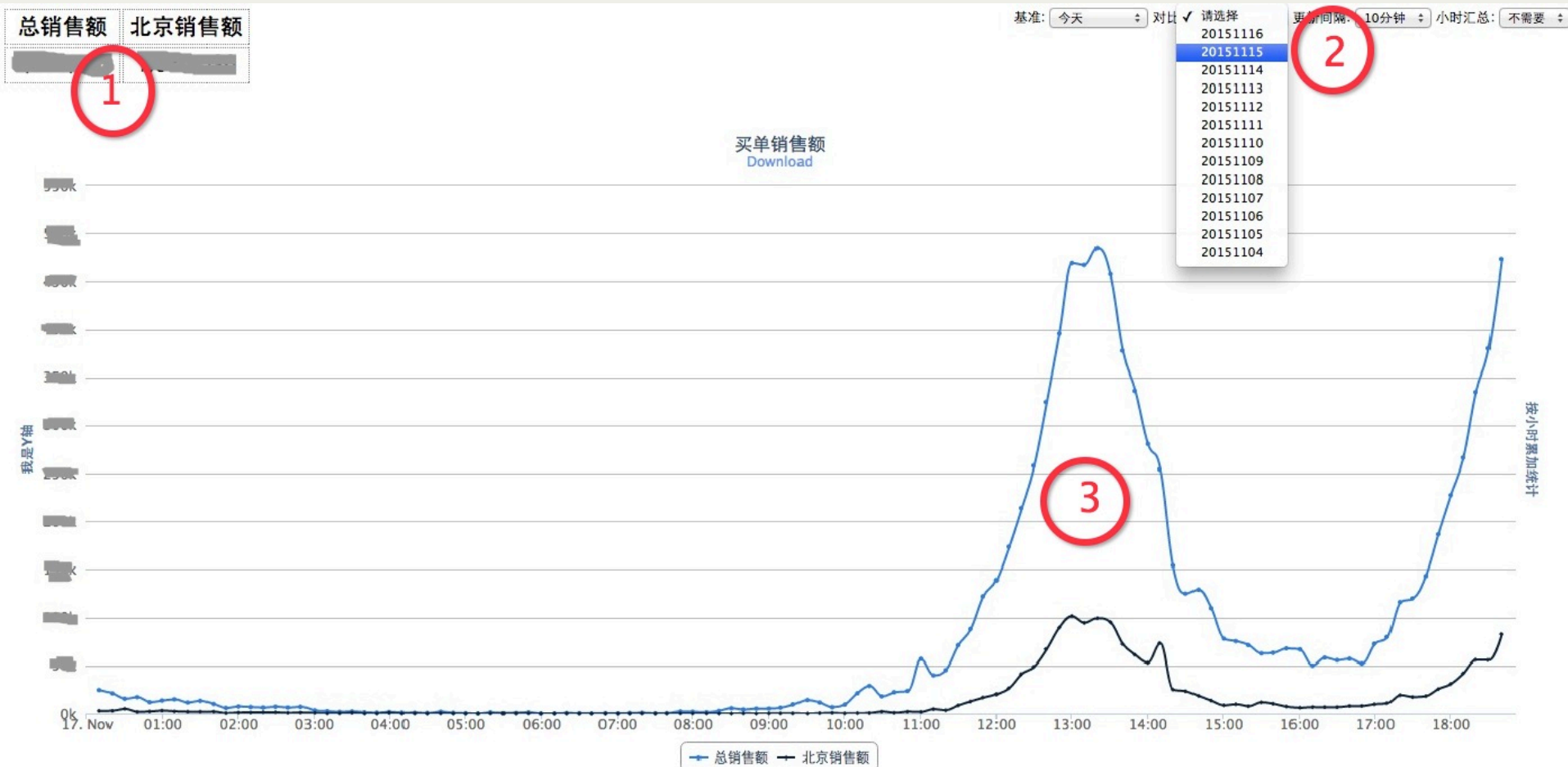
- 案例：实时数据运营
- 实时计算平台
- 实时数据仓库
- 应用：RtDW直播间
- 经验总结
- 未来规划

# RtDW直播间



- 运营和销售
- 实时指标和维度
- 定制化
- 丰富的展现

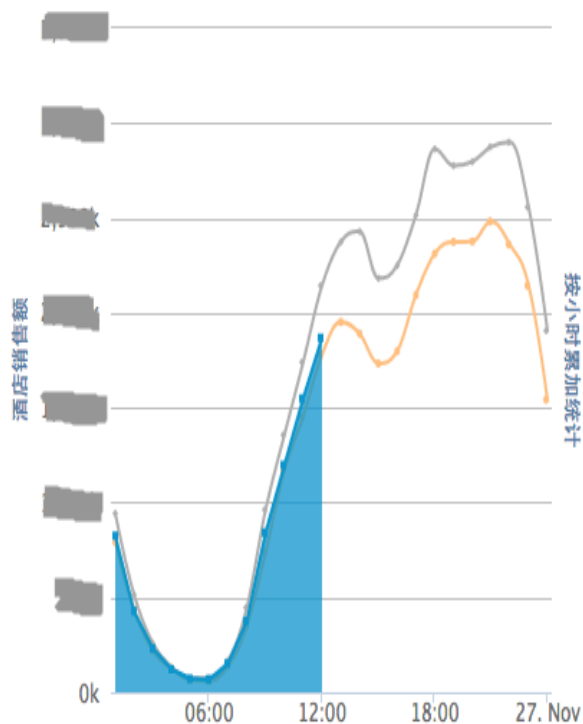
# RtDW直播间



# RtDW直播间

酒店销售额

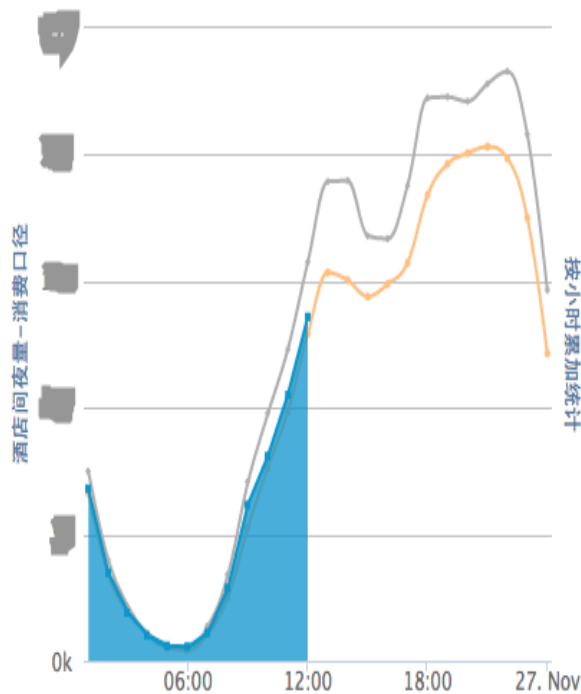
酒店销售额  
Download



— 酒店销售额\_20151125 — 酒店销售额\_20151119 ■ 酒店销售额

酒店间夜量-消费口径

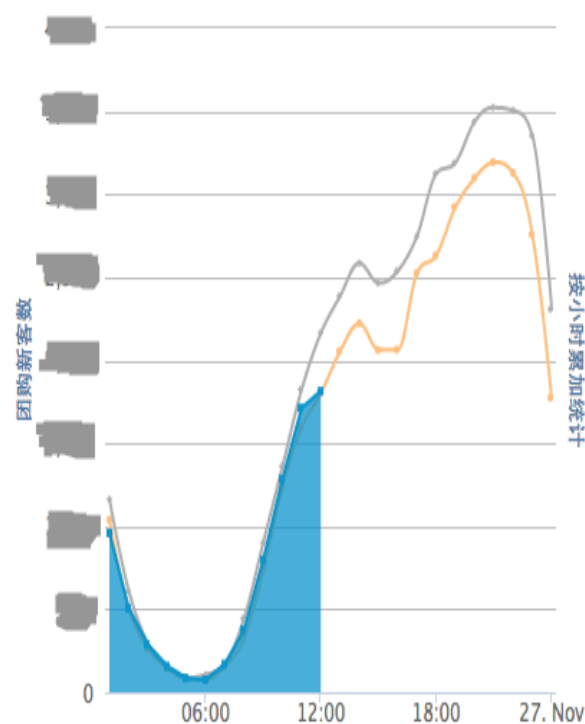
酒店间夜量-消费口径  
Download



— 酒店间夜量-消费口径\_20151125 — 酒店间夜量-消费口径\_20151119 ■ 酒店间夜量-消费口径

团购新客数

团购新客数  
Download



— 团购新客数\_20151125 — 团购新客数\_20151119 ■ 团购新客数

# 大纲

- 案例：实时数据运营
- 实时计算平台
- 实时数据仓库
- 应用：RtDW直播间
- 经验总结
- 未来规划

# 经验总结

- 仅实时化重点需求；
- Dim和State统一维护；
- 明细数据需求，基于DB，按天切表，临时聚合；
- 实时和离线统计口径的一致性；
- 历史数据的重要性；
- 修补数据的方案的重要性；

# 大纲

- 案例：实时数据运营
- 实时计算平台
- 实时数据仓库
- 应用：RtDW直播间
- 经验总结
- 未来规划



# 未来规划

- 实时和离线计算逻辑统一；
- 高可靠数据解决方案；
- 开放共建实时数据仓库；
- ...



谢谢!